ISSN: 2148-9173 **Vol: 6 Issue:3 Dec 2019**

International Journal of Environment and Geoinformatics (IJEGEO) is an international, multidisciplinary, peer reviewed, open access journal.

Generic Mapping Tools and Matplotlib Package of Python for Geospatial Data Analysis in Marine Geology

Polina LEMENKOVA

Chief in Editor

Prof. Dr. Cem Gazioğlu

Co-Editor

Prof. Dr. Dursun Zafer Şeker, Prof. Dr. Şinasi Kaya,
Prof. Dr. Ayşegül Tanık and Assist. Prof. Dr. Volkan Demir

Editorial Committee (December 2019)

Assos. Prof. Dr. Abdullah Aksu (TR), Prof. Dr. Bedri Alpar (TR), Prof. Dr. Lale Balas (TR), Prof. Dr. Levent Bat (TR), Prof. Dr. Paul Bates (UK), Prof. Dr. Bülent Bayram (TR), Prof. Dr. Luis M. Botana (ES), Prof. Dr. Nuray Çağlar (TR), Prof. Dr. Sukanta Dash (IN), Dr. Soofia T. Elias (UK), Prof. Dr. A. Evren Erginal (TR), Assoc. Prof. Dr. Cüneyt Erenoğlu (TR), Dr. Dieter Fritsch (DE), Assos. Prof. Dr. Çiğdem Göksel (TR), Prof.Dr. Lena Halounova (CZ), Prof. Dr. Manik Kalubarme (IN), Dr. Hakan Kaya (TR), Assist. Prof. Dr. Serkan Kükrer (TR), Assoc. Prof. Dr. Maged Marghany (MY), Prof. Dr. Michael Meadows (ZA), Prof. Dr. Nebiye Musaoğlu (TR), Prof. Dr. Erhan Mutlu (TR), Prof. Dr. Masafumi Nakagawa (JP), Prof. Dr. Hasan Özdemir (TR), Prof. Dr. Chryssy Potsiou (GR), Prof. Dr. Erol Sarı (TR), Prof. Dr. Maria Paradiso (IT), Prof. Dr. Petros Patias (GR), Prof. Dr. Elif Sertel (TR), Prof. Dr. Nüket Sivri (TR), Assoc. Prof. Dr. Füsun Balık Şanlı (TR), Prof. Dr. Uğur Şanlı (TR), Duygu Ülker (TR), Assoc. Prof. Dr. Oral Yağcı (TR), Prof. Dr. Seyfettin Taş (TR), Assoc. Prof. Dr. Ömer Suat Taşkın (US), Dr. İnese Varna (LV), Dr. Petra Visser (NL), Prof. Dr. Selma Ünlü (TR), Assoc. Prof. Dr. İ. Noyan Yılmaz (AU), Prof. Dr. Murat Yakar (TR), Assit. Prof. Dr. Sibel Zeki (TR)

Abstracting and Indexing: DOAJ, Index Copernicus, OAJI, Scientific Indexing Services, Journal Factor, Google Scholar, Ulrich's Periodicals Directory, WorldCat, DRJI, ResearchBib, SOBIAD

Electronic copy available at: https://ssrn.com/abstract=3504388

IJEGEO International Journal of Environment and Geoinformatics 6(2): 225-237

Reaserch Article

Generic Mapping Tools and Matplotlib Package of Python for Geospatial Data Analysis in Marine Geology

Polina Lemenkova

Affiliation: Ocean University of China, College of Marine Geo-sciences. Address: 238 Songling Road, Laoshan, 266100, Qingdao, Shandong Province, People's Republic of CHINA.

E-mail: pauline.lemenkova@gmail.com

Received 06 July 2019
Accepted 30 Sept. 2019

How to cite: Lemenkova, P. (2019). Generic Mapping Tools and Matplotlib Package of Python for Geospatial Data Analysis in Marine Geology, *International Journal of Environment and Geoinformatics (IJEGEO)*, 6(3). 225-237. DOI: 10.30897/ijegeo

Abstract

Understanding patterns of the correlation between the geomorphology and geology of the seafloor of the hadal trenches is important for the proper ocean modelling. Current paper focuses on the west Pacific Ocean region with a special case of Mariana Trench, the deepest hadal trench on the planet. Methodology of the research include combination of Generic Mapping Tools (GMT) and Quantum GIS based mapping of the geographic location, bathymetry, geodesy, sediment thickness, geomorphic shape, tectonic and geologic structure of the Mariana Trench area, and statistical analysis by means of Python. A GMT was selected for GIS visualization due to its powerful functionality and effective cartographic solutions. An object-oriented high-level programming language, Python was chosen for the data analysis and scientific plotting. The statistical analysis includes following steps: 1) Data distribution by the box plots; 2) Data sorting and grouping by stem plots; 3) Correlation analysis by 3D comparative plots referred to four tectonic plates; 4) Principal Component Analysis; 5) Analysis of Variance. The statistical analysis of the data set was performed in Matplotlib library and its dependencies: NumPy, SciPy and Pandas. A combination of the powerful methods by GMT with data analysis supported by Python programming language is an important method in geosciences aimed to increase the effectiveness of the data analysis by cartographic mapping, statistical computations and graph plotting. This paper illustrated usage of GMT, QGIS and Python for combined data analysis scheme. The results demonstrated correlation between the sediment thickness, slope steepness, depths and location of the bathymetric profiles crossing adjacent tectonic plates: Philippine, Pacific, Caroline and Mariana.

Keywords: GMT, Mapping, Python, Statistics, Mariana Trench, Marine Geology.

Introduction

The study object is Mariana Trench, located in the west Pacific Ocean (Fig. 1). Mariana Trench is formed in the area of subduction of the tectonic plates, Pacific Plate subducts beneath the Philippine Sea Plate, Mariana and Caroline Plates are located besides. Mariana Trench is the deepest hadal trench on the Earth having the oldest oceanic crust. Understanding patterns of the correlation between the geomorphic and geological features of the seafloor of such unique hadal trench is critical for its proper modelling. There are multiple factors that affect the structure and formation of the trench: geology, geographic location, geodetic properties, gravimetry, tectonic plates, faults and slabs, location of large igneous provinces, ophiolites (Fig. 2). For instance, correct modelling of the accumulation of the sediment thickness with reference to the trench location, its underlying tectonics and geology may visualize interrelationships between the deep-sea deposits and environmental changes at the sea surface. Large geological data sets are especially well suited to the quantifying the correlations between the seafloor and other environmental factors.

Geospatial analysis in its broadest application covers the GIS based mapping techniques and statistical analysis allowing to detect correlations, trends and behavior of the spatial variables in space. Applied to the marine geological studies, a geospatial analysis incorporates not only the statistical and geoinformation techniques, which are the main focus of this research, but also satellite observation data covering ocean regions, cruise ship based geophysical surveys aimed to measure complex and often subtle changes in the submarine landform surface and precise bathymetry.

Whatever the technique employed, the central focus is laid on the geospatial data analysis in its times series or spatial series presentation form. To analyze geomorphological landform features by means of the statistical analysis with GIS data combination, the main question is how the landform change over the space and what are the reasons for variability in the geomorphology. The impact factors that affect topography can be categorized into those with direct effects, that is, tectonics, geological features, geographic location, and indirect effects, e.g. inducing local effects on the landforms, for example accumulation of the biomass and sedimentation rates (Craglia et al., 2012-2017; Jiang, et al., 2019).

The geomorphological variations in the submarine landforms of the Mariana Trench is determined by several environmental and geological variables, such as

Electronic copy available at: https://ssrn.com/abstract=3504388

location of faults, ridge lines and geographic stretch of the trench arc, patterns of the geological structure of the seafloor. Location of the large igneous provinces affects the volcanism activities in the nearby area. Pattern, intensity and direction of the ocean currents determine the up-bring the nutrients to the deep ecosystems. Apart from geology and geographic settings, location of the tectonic plates, slab movements direction is another important factor of the possible changes in the slope geomorphology of the trench. The combination of these factors alters fundamental processes that have geomorphic consequences for the shape of the hadal trench. Understanding the scope and direction of the geomorphic change caused by geologic and tectonic factors requires a multi-disciplinary approach of data processing, analysis and modelling by GIS and statistical tools. This is particularly true for such remotely located submarine geomorphic systems as Mariana Trench. This study aims at geospatial analysis and statistical data modelling of the variations of geomorphology of the Mariana Trench, and investigates the correlation of data with trench geomorphology.

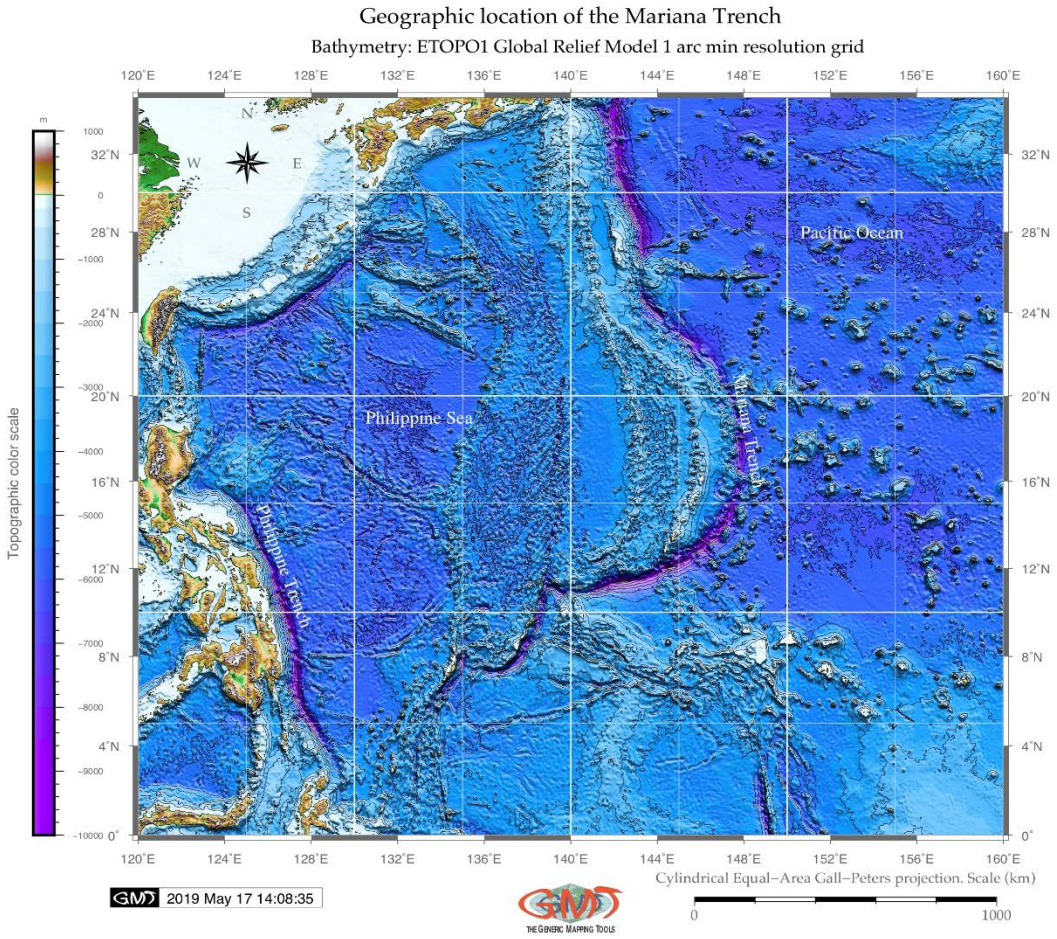

Fig. 1. Geographic location of the Mariana Trench

Besides practical geographic purpose of geomorphic studies, geospatial marine data modelling is necessary for the robust estimation of the ocean geological mineral resources. The reconstruction and prediction of the sea-floor resources and deposits for mining, as well as estimation of the climate change and its impact on the ocean environment requires precise computation of the geological data sets. Visualized main geological features of the study region include stretching of the Mariana Trench, location of ridges, transform lines, distribution of the ophiolites and large igneous provinces (Fig. 2). The precise geoid determination is crucial for various tasks of the geospatial analysis such as determination of physical heights, establishment of a height control in remote areas, precise mapping, connection of different height system, etc. Recent advances in the development of the gravity field satellite missions, gravimetric geoid determination and progress in the algorithms of big data processing made it possible to increase the precision of the geodetic measurements. Furthermore, accurate determination of the geoid using satellite gravity models is now available through the CryoSat-2 and Jason-2 satellite missions together with high-resolution (1 minute) terrestrial gravity and topographic data (Fig. 4 and Fig. 5). Accurate geoid modelling enables to provide a continuous equipotential reference surface that is a natural vertical datum for heights with the vertical reference surface based on a gravimetric geoid with height datum information. There are existing approaches quantifying and modeling the geoid and dynamic topography by density anomaly distributions using from seismic tomography (Ricard et al., 1989; King & Masters, 1992). Geoid modelling in the Mariana Trench area (Fig. 3) was performed using World Geoid Image version 9.2, 2 minute resolution, 72S to 72N (Sandwell & Smith, 1997; Smith & Sandwell, 1997).

Electronic copy available at: https://ssrn.com/abstract=3504388

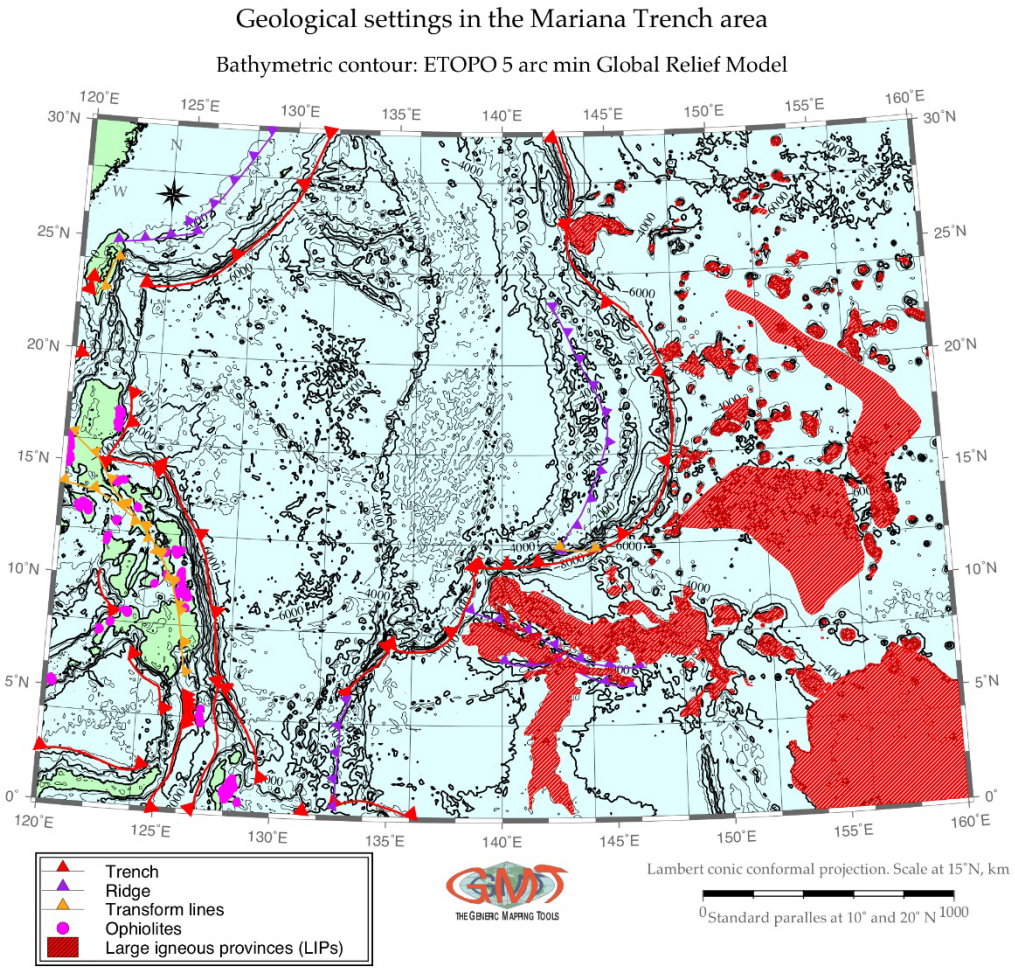

Fig. 2. Geological settings of the study region.

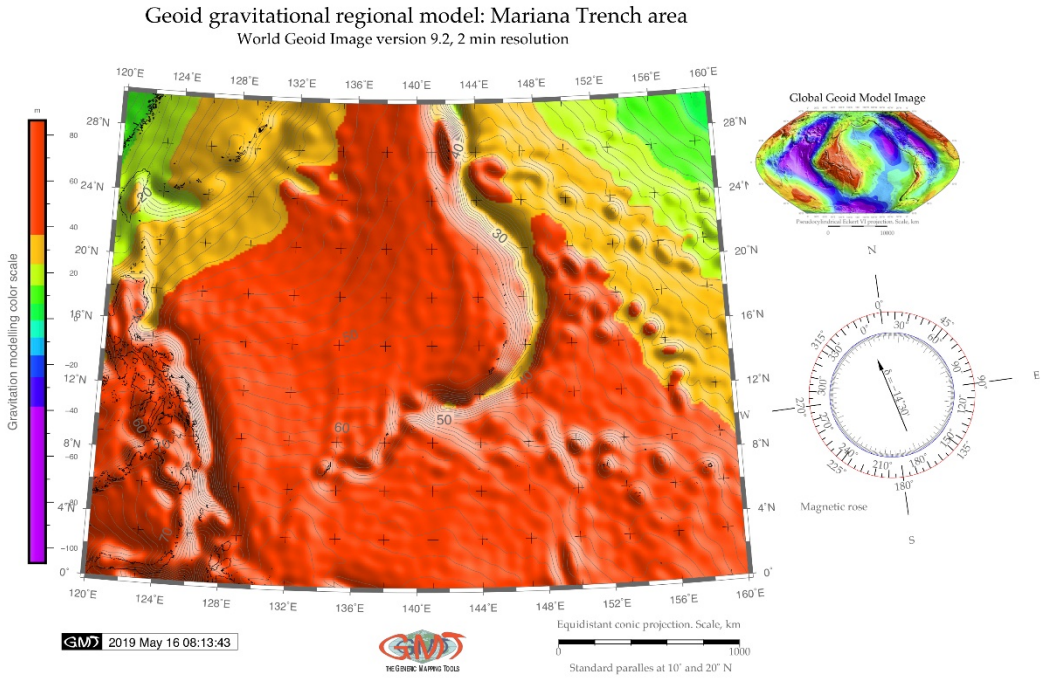
Geoid gravitational regional model: Mariana Trench area
World Geoid Image version 9.2, 2 min resolution
Gravitation modelling color scale
Global Geoid Model Image
Magnetic rose
2019 May 16 08:13:43
THE GENERIC MAPPING TOOLS
Equidistant conic projection. Scale, km
Standard parallels at 10° and 20° N
0
1000

Fig. 3. Geoid modelling of the study area

Electronic copy available at: https://ssrn.com/abstract=3504388

Conversion of the tectonic plates at the ocean margins causes sinks or subduction of the oceanic plates. The mechanism of faulting in the subduction plates at the Mariana Trench was described previously (Zhou et al., 2015; Zheng et al., 2013). The terrestrial crust appearing at the top of this new lays is formed either by a new masses of the terrestrial crust caused by the arc volcanism, or because of the piled up sediment deposits coming from the subducting lower tectonic plate (Heune & Scholl, 1991).

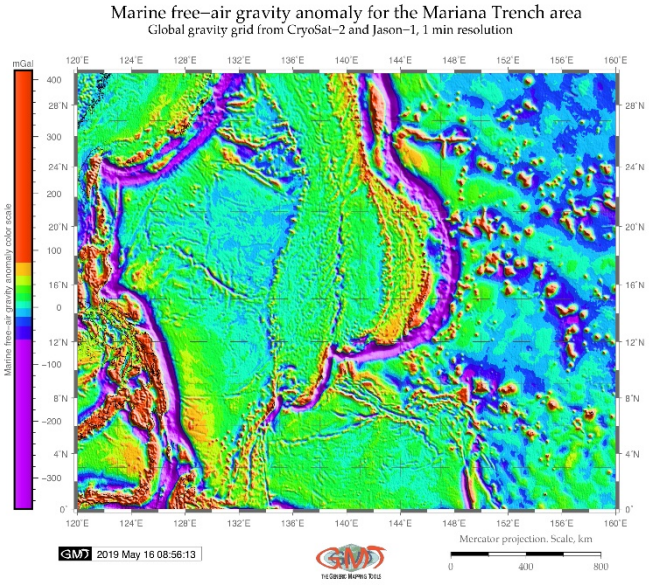
Marine free−air gravity anomaly for the Mariana Trench area
Global gravity grid from CryoSat−2 and Jason−1, 1 min resolution
Marine free−air gravity anomaly color scale
Mercator projection. Scale, km
GMT 2019 May 16 08:56:13

Fig. 4. Gravitation model of the study area

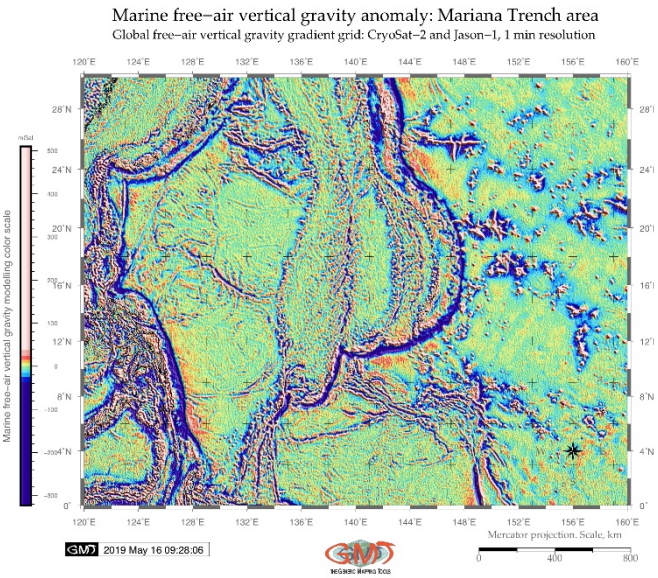
Marine free−air vertical gravity anomaly: Mariana Trench area
Global free−air vertical gravity gradient grid: CryoSat−2 and Jason−1, 1 min resolution
Marine free−air vertical gravity modelling color scale
Mercator projection. Scale, km
GMT 2019 May 16 09:28:06

Fig. 5. Vertical free-air gravitation model of the study area

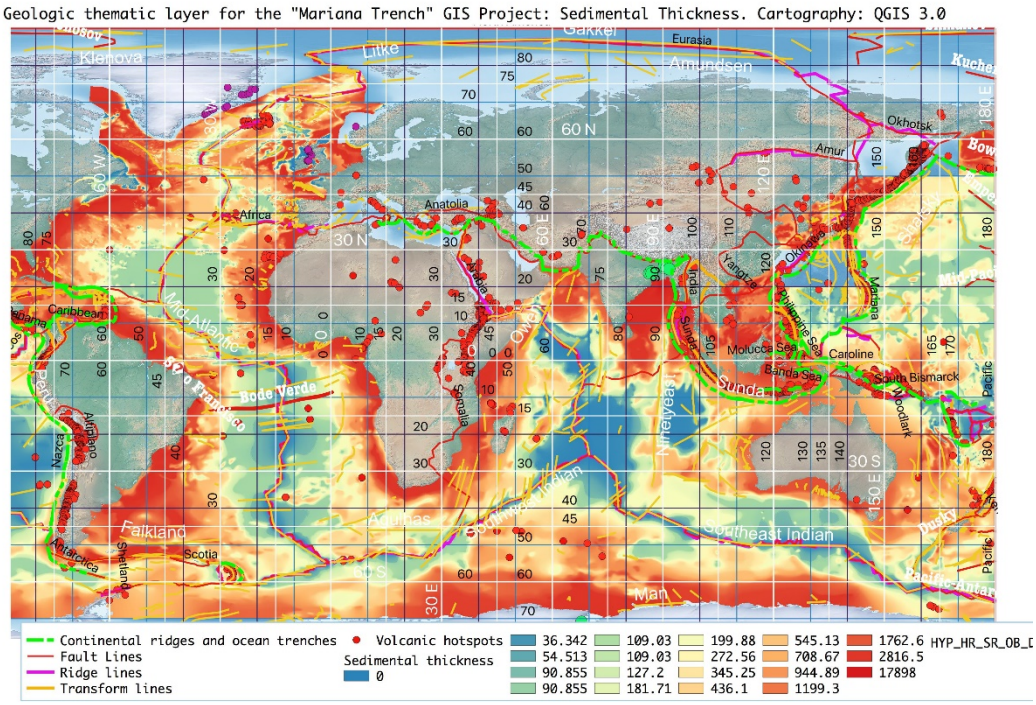

Fig. 6. Sediment thickness in the World Ocean. Mapping: QGIS

Therefore, the ocean margin areas with high seismicity, repeated earthquakes and volcanism indicate the areas of the collision and movements of the ocean plates as well as active sedimentation, where hadal trenches are located. Such and other geodynamic processes of the deep subduction take place of the western Pacific Plate underneath Northeast China causing extensional events and volcanism (Zhu et al., 2010). Specific changes in the geochemical content of the underlying seafloor has also beed detected. Consequences of the tectonic movements create changes in the geochemical properties of the seafloor. Thus, as recently noticed by Cartapanis et al. (2018), the content of the geological CO2 burial rates varies in the seafloor, with the increase in the Pacific basin during the last glacial period. There are various two distinctly different domains of the seafloor exist, on which carbon burial takes place: surfaces located by the fringes of the continents and the constantly submerged surfaces of continental slopes and the deep sea. The

Electronic copy available at: https://ssrn.com/abstract=3504388

mixture of the geochemical components located at various depths of the continental slope and trench implies that sediments from upper to lower slope has been incorporated into a single bed (Völker et al., 2008). Other examples of the geochemical complex circulation, system of deposition and transport given by various studies (e.g., Sattarova & Aksentov, 2018; Yang et al., 2018; Lehu et al., 2015).

Oceanic lithosphere has a complex structure varying in different parts of the margin seas located in the areas of the subducted tectonic plates. General correlation between the subduction earthquakes, sediment thickness on the seafloor of the trench and upper plate strain has been well discussed and proved by Heuret et al., 2012. Specifically, the margin trenches can illustrate the correlation of the sedimentation with the location of the trench where sediment accumulation results in a significant load (Fig. 6). Thick trench fill sediments may lead to the obscured morphologic expression of the outer part of the trench or deformed front. Sediment accumulation affect the lithospheric downward deflection near the trench and the amount of the subducted sediment (Contreras-Reyes et al., 2013). Moreover, the effective elastic thickness at the subduction zones of the tectonic plates has important impacts on the tectonic and geodynamic processes of the hadal trench, providing restrictions for the strength of the oceanic lithosphere (Yang & Fu, 2018).

Besides geologic factors, there is an interconnection between the sediment thickness of the trench and the biological life in the hadal trenches: scarcity of food is one of the factors limiting the presence and functioning of the heterotrophic organisms at hadal trenches, besides the high hydrostatic pressure. In turn, the availability of the nutrients largely relies on the flux of the organic-rich matter depositing from the ocean surface (Turnewitsch et al., 2014). There are many published reports on the typology and classification of the marine sediments on the ocean floor (e.g., Lisicyn, 1974; Udintsev, 1987; Gazioğlu et al., 2002-2005; Dolotov, 2010; Dutkiewicz et al., 2015; Dong et al., 2018; Lacharité & Metaxas, 2017). The approaches used to generalized into using standard geospatial analysis aimed at description of the main problems of the location and features of the deposited sediments in various parts of the global ocean (Fig. 6).

Materials and Methods

The aim of this study is to perform geospatial analysis of the cross-sectional profiles of the Mariana Trench: to determine what are the most/least deep profiles, what are the most/least steep slope gradient angles by these profiles, in which segment of the Mariana Trench arc they are located, how environmental factors correlate with each other by four tectonic plates.

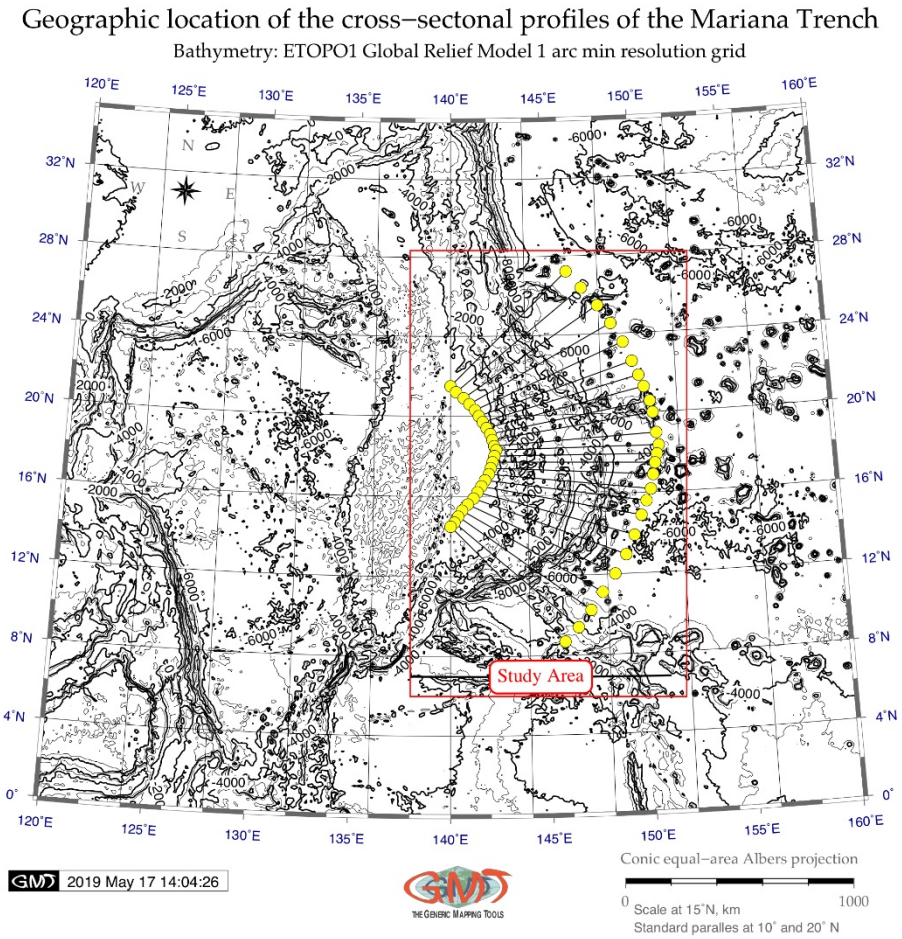
Geographic location of the cross−sectonal profiles of the Mariana Trench
Bathymetry: ETOPO1 Global Relief Model 1 arc min resolution grid
Study Area
2019 May 17 14:04:26
Conic equal−area Albers projection
Scale at 15°N, km
Standard paralles at 10° and 20° N

Fig. 7. Location of the cross-sectioning 25 profiles

Electronic copy available at: https://ssrn.com/abstract=3504388

The research methodology encompasses comparing attribute data received in the points of the intersection of the by 25 cross-section track profiles (Fig. 7) with thematic GIS layers. In addition, several thematic maps on the study area were plotted (Fig. s 1-6) using GMT and Quantum GIS software.

Python libraries

The research approach consists in the processing geodata received from the attribute tables of the geospatial layers. The cartographic part of this research included plotting thematic maps from the grid data sets (Fig. 1-6) and vectorized cross-sectioning 25 profiles (Fig. 7) in Generic Mapping Tools (GMT) environment (Wessel & Smith, 1995). Mapping using Generic Mapping Tools (GMT) was performed using available documentation on cartographic workflow (Wessel et al., 2017) following the GMT modules: img2grd, grd2cpt, grdimage, pscoast, psbasemap, psscale, logo, pstext, grdcut, grdinfo, grdcontour, psbasemap, psxy. Vector layers were stored in the Quantum GIS (QGIS) projects. The data were then imported into the tables with a csv extension for further statistical processing using Python.

The broad application of the Python language for marine geological modelling presented in current paper is based on the application of the embedded mathematical algorithms and libraries of Python: Matplotlib, NumPy, SciPy, Statsmodel. The syntax of the written Python's codes include operating with both immutable objects (such as strings and tuples) and mutable sequences of mixed types of items (such as lists). Besides, the division of the code into a visible interface and run makes the coding intuitive and fast. Moreover, the important feature of the Python consists in its object-oriented approach: encapsulating the long lists of the arguments and passing them into the objects makes the implementation of the code effective and not overloading the computer memory, which is freed shortly after the last reference to it has been removed. Finally, syntax coding in Python can be formatted in readable and very structured style due to the long history of the language development. At the first step, the necessary statistical Python libraries were installed. These include Matplotlib, NumPy and SciPy. The SciPy library has many modules used for numerical computations. It is used for the statistical data transformation, numerical simulation, modeling, visualization by machine learning algorithms. The NumPy library was installed for reading tables from the .csv format that were derived from the QGIS database from the attributive tables of vector layers. The Matplotlib library is the main library used for statistical data processing, visualization and plotting. The installation was done using command via bash terminal (example for Matplotlib): $ python3 -m pip install matplotlib.

The next step included installing of the certificates to enable connections to the websites containing data frames in Python 3.7.1 on the Mac OS Mojave 10.14.3. This step was necessary, because the Python 3.x has no certificates. This action required a post-install step installing the 'certifi' package, version certifi-2018.11.29. To validate SSL connections the following command has been implemented in the bash terminal:

- $ sudo /Applications/Python\ 3.7/Install\ Certificates.command

The Jupyter Notebook, an open-source web application, was installed using bash command:

- $ sudo python3 -m pip install jupyter

It is used to store and share Python codes, equations and plot visualizations with legends. Running Jupyter Notebook was done using the call: $ jupyter notebook. All the scripts and codes were processed, saved and exported via the Jupyter Notebook.

At the third step the data were taken in a table form (.csv) from the attribute tables of the vector layers obtained from the GIS project operated by the QGIS and all further data processing was done using Python. Python plotted all the graphics and illustrations presented in this research. The Matplotlib package was selected as a main tool, as it enables operating statistical analysis and plot Fig. s using 'pyplot' functions: plotting and adjusting the areas, visualizing statistical outputs, selecting colors for the plots, adding labels etc.

Data distribution

The box plots (Fig. 8) were plotted to demonstrate the data distribution across the 25 profiles with vertical lines extending from the boxes (so called whiskers) indicating variability of the depths outside the upper and lower quartiles. The plot was made via the following Python code (1) and supported by the existing methods of the statistical analysis (Frigge et al., 1989; Bates & Watts, 1988; McGill et al., 1978).

Python code (1) for a box-and-whisker plot for the Mariana Trench bathymetry

```
import seaborn as sb
from matplotlib import pyplot as plt
import pandas as pd
import os
os.chdir('/Users/pauline/Documents/Python')
dfB = pd.read_csv("Tab-Bathy.csv")
dfB.head(5)
sb.set_style('whitegrid')
sb.boxplot(data=dfB, orient="v", palette='coolwarm', saturation=1,
    width=0.8, dodge=True, fliersize=5, linewidth=0.2,
    whis=5, notch=True, ax=None)
sb.despine(offset=10, trim=True) # to offset the spines away from the bathymetric data
plt.xticks(rotation=45)
plt.yticks(rotation=45)
plt.title('Box-and-whisker plot for the Mariana Trench bathymetry', fontsize=12, fontfamily='sans-serif')
plt.show()
```


Electronic copy available at: https://ssrn.com/abstract=3504388

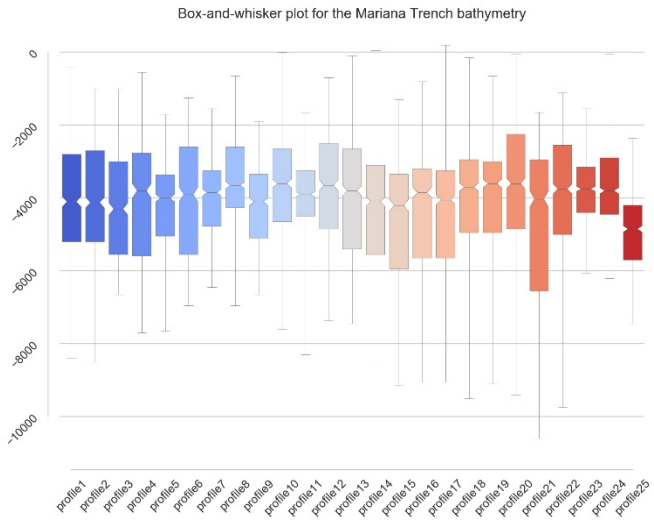

Fig. 8. Box-and-whisker plot for the bathymetric depths distribution by 25 profile

Data sorting and grouping

Grouping profiles according to the slope steepness (Fig. 9) was aimed at visualizing the most steep parts of the trench and identifying the steepest profiles and, on the contrary, those with gentle slopes. Initially, the profiles were plotted in a geographic order (Fig. 9 A), as they follow each other from north in southward direction (Fig. 7). The second step included ranking profiles according to their steepness. The steps representing the profiles are re-labelled accordingly. The modelling was performed using following code in six steps.

Python code (2) for sorting data in stem plots: analysis of the angle steepness:

- # Step-1. loading libraries
- import matplotlib.pyplot as plt
- import numpy as np
- import pandas as pd
- import os
- import pylab as plt
- # Step-2. defining Fig.
- fig = plt.Fig. ()
- st = fig.suptitle('Geomorphologycal analysis of the Mariana Trench: 25 bathymetric profiles \nunsorted (A) and sorted (B) plots by steepness angles', fontsize=12)
- ax1 = fig.add_subplot(211)
- ax2 = fig.add_subplot(212)
- # Step-3. defining subplot 1
- plt.subplot(2, 1, 1)
- os.chdir('/Users/pauline/Documents/Python')
- df = pd.read_csv("Tab-Morph.csv")
- x=df.profile
- values=df.tg_angle
- markerline, stemlines, baseline = plt.stem(x, values, markerfmt="o", basefmt='-')
- plt.setp(markerline, mfc='salmon', markersize=13, alpha=0.7, markeredgewidth=.2)
- plt.setp(stemlines, linewidth=.5, color='dimgray')
- plt.setp(baseline, color='purple', linewidth=1)
- plt.xlabel('Bathymetric profiles', fontsize=10, fontfamily='sans-serif')
- plt.ylabel('tg$^\\circ$ (A/H)', fontsize=10, fontfamily='sans-serif')
- # Step-4. defining subplot 2
- plt.subplot(2, 1, 2)
- os.chdir('/Users/pauline/Documents/Python')
- df = pd.read_csv("Tab-Morph.csv")
- df = pd.DataFrame({'group':df.profile, 'values':df.tg_angle}) # take only two values: bathymetric profiles and angles
- ordered_df = df.sort_values(by='values') # sort df by values
- markerline, stemlines, baseline = plt.stem(x, ordered_df['values'], markerfmt="o", basefmt='-')
- plt.setp(markerline, mfc='salmon', markersize=13, alpha=0.7, markeredgewidth=.2)
- plt.xticks(x, ordered_df['group'])
- plt.setp(stemlines, linewidth=.5, color='dimgray')
- plt.setp(baseline, color='purple', linewidth=1)
- plt.xlabel('Bathymetric profiles', fontsize=10, fontfamily='sans-serif')
- plt.ylabel('tg$^\\circ$ (A/H)', fontsize=10, fontfamily='sans-serif')
- # Step-5. annotating labels
- ax1.annotate('(A)', xy=(1.02, .90), xycoords="axes fraction")
- ax2.annotate('(B)', xy=(1.02, .90), xycoords="axes fraction")
- # 0,0 is lower left of axes and 1,1 is upper right
- # Step-6. plotting
- plt.subplots_adjust(hspace=.2)
- plt.show()

Electronic copy available at: https://ssrn.com/abstract=3504388

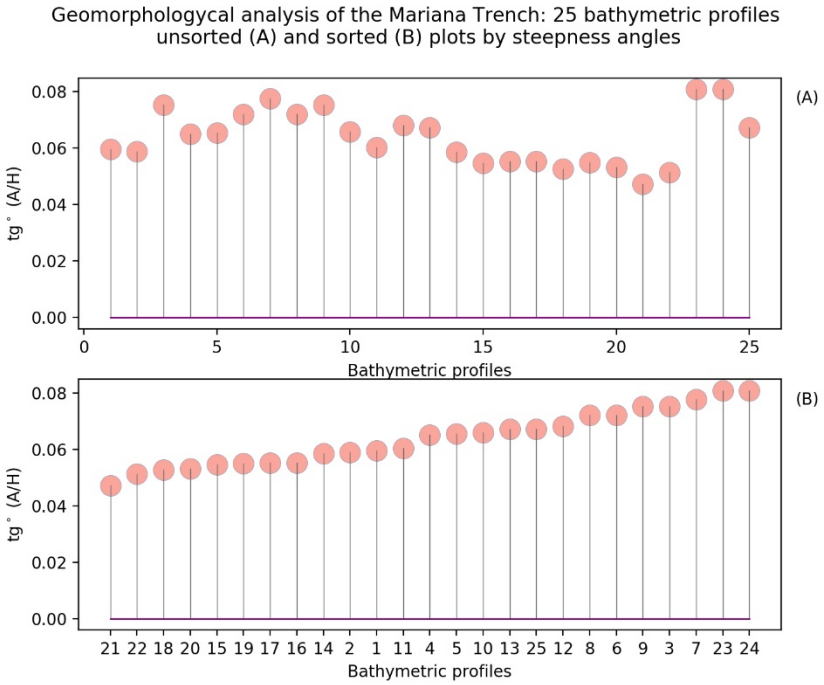
Geomorphologycal analysis of the Mariana Trench: 25 bathymetric profiles
unsorted (A) and sorted (B) plots by steepness angles
tg ° (A/H)
Bathymetric profiles
(A)
(B)
21 22 18 20 15 19 17 16 14 2 1 11 4 5 10 13 25 12 8 6 9 3 7 23 24

Fig. 9. Stem plot of the geomorphic analysis of the slope steepness by 25 profiles.

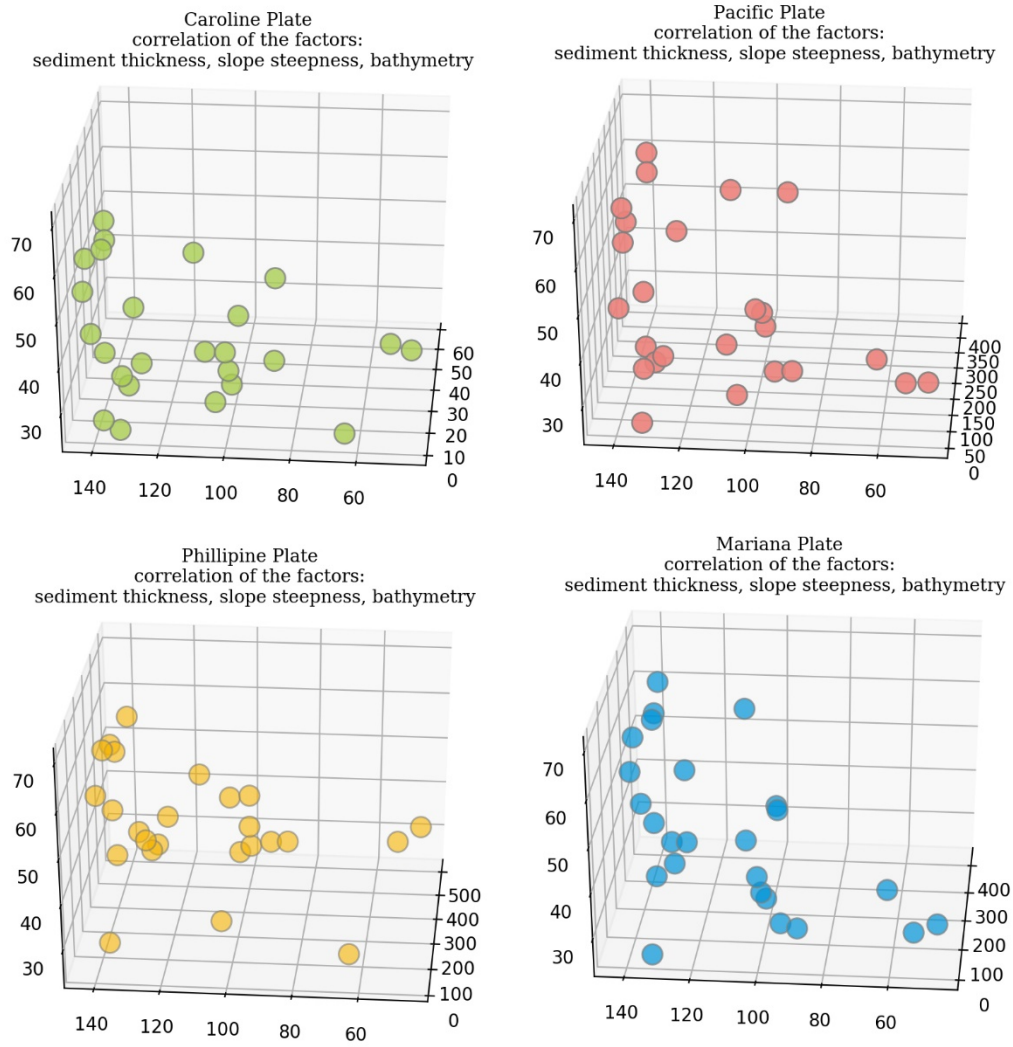

Fig. 10. 3D Scatterplot of the interrelationships between the environmental factors

Electronic copy available at: https://ssrn.com/abstract=3504388

Data correlation

The analysis data by correlation of the environmental and geological factors by four tectonic plates (Fig. 10) was performed using 3D scatterplot made by means of the ax.scatter() function of the Matplotlib library using following code in 3 steps (Code 4):

Python code (3) for 3D data plotting:

- # Step-1. libraries
- from mpl_toolkits.mplot3d import Axes3D
- import matplotlib.pyplot as plt
- from matplotlib.colors import ListedColormap
- import numpy as np
- import pandas as pd
- import matplotlib.cm as cmx
- # Step-2. Dataset
- df = pd.read_csv("Tab-Morph.csv")
- # Step-3. Plotting
- fig = plt.Fig. ()
- ax = fig.add_subplot(111, projection='3d')
- ax.scatter(df.plate_maria, df.sedim_thick, df.slope_angle, c='#0095d9', edgecolors='grey', s=130, alpha=.7)
- ax.view_init(30, 185)
- plt.title('Mariana Plate \ncorrelation of the factors: \n sediment thickness, slope steepness, bathymetry', fontsize=10, fontfamily='serif')
- plt.show()

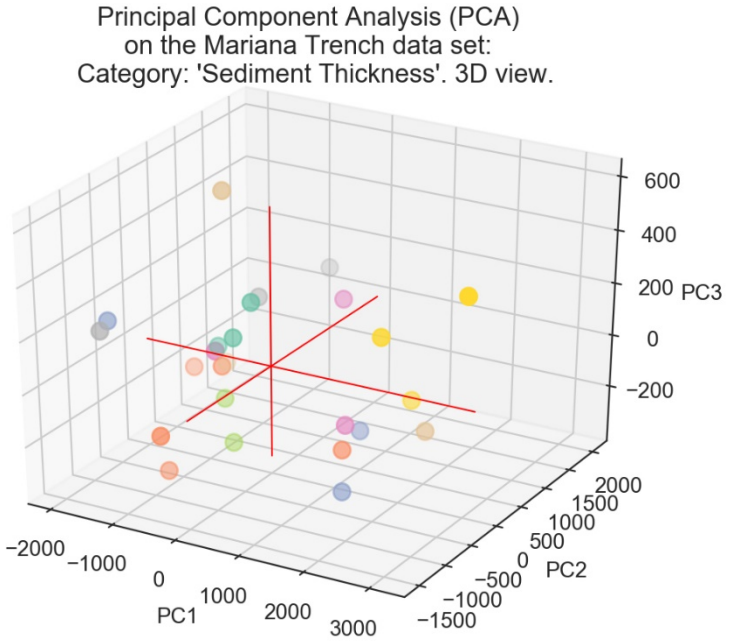

Fig. 11. PCA of sediment thickness by 25 bathymetric profiles, Mariana Trench. 3D view

Principal Component Analysis (PCA)

The Principal Component Aanalysis (PCA) assumes computing the statistical importance of the mean and covariance within the data. The PCA modelling was made using the 'PCA' function from Sci-Kit Learn library of Python (from sklearn.decomposition import PCA). Using orthogonal transformation procedure embedded in PCA to convert a set of observations on sediment thickness and variables and possibly correlated variables into a set of values of linearly uncorrelated variables (Kaya et al., 2004; Abdi & Williams, 2010). Using PCA is an especially useful technique in case of marine geological large datasets and other implications of the big data in the geosciences, where there are many variables for each sample for exploratory data analysis (Jolliffe, 2002). The PCA enable to better visualize (Fig. 11) the variation presented in a dataset with many variables, that is sediment thickness, steepness of the slope angles by cross-secting profiles and bathymetric depth ranges.

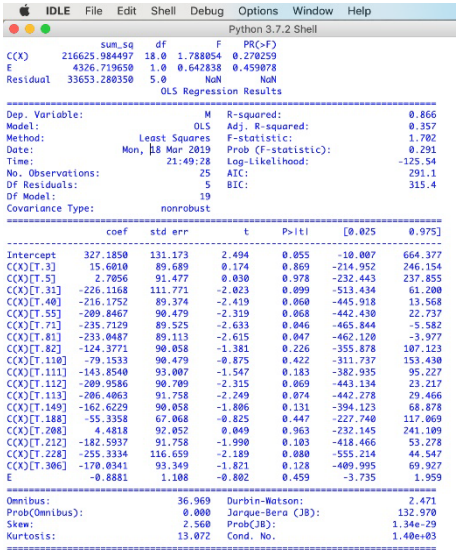

Fig. 12. ANOVA table computed by Statsmodel library and visualized in Python shell

Electronic copy available at: https://ssrn.com/abstract=3504388

Analysis of Variance (ANOVA)

The purpose of the Analysis of Variance (ANOVA) was to test significant differences between the means of the marine environmental and geological variables, that is geological, tectonic and geodetic factors (Fig. 12 and Fig. 13). Test differences in means for groups of mentioned above variables for the statistical significance was done using existing algorithms (Gelman, 2005; Montgomery, 2001; Bailey, 2008).

The ANOVA method is accomplished by analyzing the variance, that is, by partitioning the total variance into the component that is due to true random error (i.e., within-group SS) and the components that are due to differences between the means. The ANOVA was performed using Statsmodels library of Python and Pandas for table data processing using Python code (5). The analysis of variance was done using Statsmodel library of Python and embedded function 'ANOVA' by the following Python code (5):

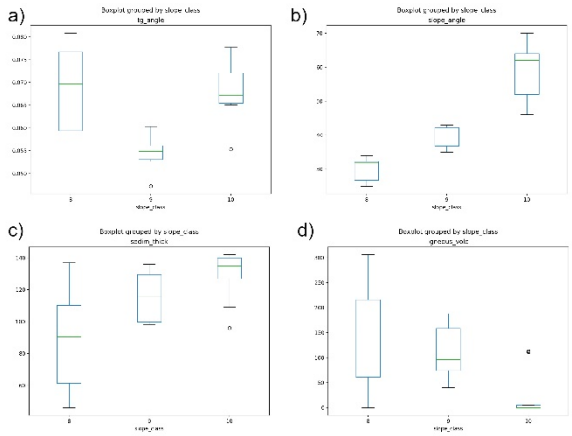

Fig. 13. One-way ANOVA box plots showing a potential relationship between selected environmental, bathymetric and geomorphological variables by analysis of similarities

Python code (4) for ANOVA:

```
# Libraries
from __future__ import print_function
import scipy
import os
from statsmodels.compat import urlopen
import numpy as np
import pandas as pd
pd.set_option("display.width", 100)
import matplotlib.pyplot as plt
from matplotlib.patches import Polygon
import statsmodels.api as sm
from statsmodels.formula.api import ols
from statsmodels.graphics.api import interaction_plot, abline_plot
from statsmodels.stats.anova import anova_lm
os.chdir('/Users/pauline/Documents/Python')
df = pd.read_csv("Tab-Morph.csv")
E = df.sedim_thick
M = df.plate_pacif
X = df.igneous_volc
S = df.slope_angle
cw_lm=ols('M ~ S + C(E)', data=df).fit()
print(sm.stats.anova_lm(cw_lm, typ=2))
print(cw_lm.summary())
cw_lm=ols('M ~ E + C(X)', data=df).fit()
print(sm.stats.anova_lm(cw_lm, typ=2))
print(cw_lm.summary())
```

Results

Together with depth, the geomorphic shape of the Mariana Trench plays a crucial role in determining patterns of the sediment thickness in the proposed model. The depth ranges increased along the axis of the Mariana Trench, from north to the south of the trench, with a relatively constant slope gradient, until reaching depths overstepping 11.00 m (profiles 20 and 21 as shown on the Fig. 8). In the south-western part of the trench axis the modeled correlation between the factors was relatively high and yet decreased towards the south-western part of the profile. A geomorphological analysis was performed (Fig. 9), using only the calculated values of the slope steepness degree by profiles in the series of 25 cross-section bathymetric profiles (assuming constant values of the slope aspect), using bathymetric values and observation points as a reference. This analysis shows that the (profiles Nr. 6, 9, 3, 7, 23 and 24) have the similar steepness and can be grouped as the lest steep, as demonstrated by the reverse values of the tangent angles. Conversely, the profiles Nr. 20, 21, 22, and 18 have the deepest values that can be explained by the changes in the geometry of the trench crossing different tectonic plates. The 3D scatterplot matrices (Fig. 10) show variability of the geological, geometric and bathymetric factors (that is, sediment thickness, steepness of slope degrees and depth values) that is similar to the model by four tectonic plates: Pacific, Mariana, Caroline and Philippine. The Principal Component Analysis (PCA) on the bathymetric spatial series was performed in order to assess the overall spatial geomorphic variability of the cross-section profiles and their similarities (Fig. 11). For this calculation, the sediment thickness as a categorical value of the studied area was used. The deep-sea sediments has consistent correlation with the slope class of the profile steepness (Fig. 13, c). Lower effect has the location of the igneous volcanic areas nearby the trench on the geometry of the profiles, as referred by the angle degree (Fig. 13, d). A visual word cloud of the approaches and methods used in this research is shown in Fig. 14 (word cloud in a fish form) to visually illustrate the main research concepts of the current work.

The sediment thickness is broadly correlated with the slope steepness degree, showing consistent geographic patterns by two pairs of plates: firstly, Mariana and Pacific, and secondly, Caroline and Philippine. Computed F values (0.270259 and 0.459078) and statistical significance obtained from testing with ANOVA shows the effect of depth and model scenario on the sediment thickness estimated through the Statsmodel library of Python (statsmodels.stats.anova anova_lm). While the shape of the trench is less steep and the absolute depths are lower in the upper part of the

Electronic copy available at: https://ssrn.com/abstract=3504388

trench, that is, north-east while crossing the Pacific and Mariana tectonic plates, the bathymetry patters is substantially deeper in the south-eastern part of the Mariana Trench where it is crossing Caroline and Philippine Sea Plates. These findings are also supporting the results of Zhou & Lin (2018) who studied variations in the deformation of the subduction plate along the Mariana Trench by analyzing its flexural bending and normal fault features. They reported that most normal faults initiated outwards and grew toward the trench axis and the trench relief and maximum fault throws are significantly greater in the south (5 km and 320 m) comparing to the northern and central areas (2 km and 200 m).

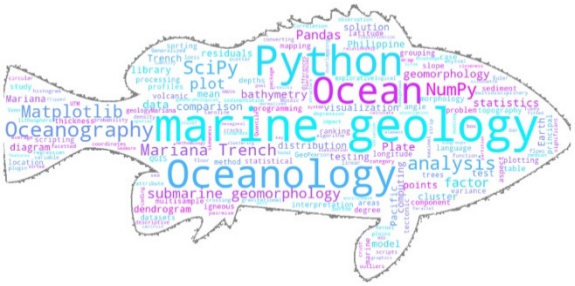

Fig. 14. Cloud 'fish'-word summarizing the methods, tools and concepts of this research.

Discussion and Conclusion

The big data are often used in the marine research, because the data are taken from the cruise observations and measurements in in-situ conditions. Therefore, the advantages of the Python language lies in its usage for marine sciences lie in its effectiveness and functionality when operating large data sets. The machine learning tools are highly effective for processing arrays and tables containing data of thousands of observation, for example as in the presented research where the data set was 12.590 points taken from the 25 profiles containing 518 data each. Python is an interpreted language that enables programs to be written compactly and readably splitting them into reusable modules (Rossum et al., 2018) that in turn process and model big data sets. Simple and vectorized syntax structures of Python enable to write code concisely yet effectively. Using NumPy for processing tables in .csv format and arrays facilitate advanced mathematical operations on large numbers of data. The encapsulation, inheritance and polymorphism of Python are the main advantages while working with big data in marine science and processing large tables.

There are proven correlations between various geodetic features, bathymetric locations of the ocean trenches, geomorphic shapes and geophysical properties of the submarine phenomena (for example, Mckenzie & Bowin, 1976; Sandwell & Smith, 2009; Jimenez-Diaz et al., 2014; Judge & McNutt, 1991; Kalnins & Watts, 2009). Understanding the deep interconnections between various, part of the ocean ecosystem gives a key to the proper modelling of the location of the ocean mineral resources and prognosis of the dynamic processes, such as earthquakes. The importance of the precision of such numerical computing for the resource economies cannot be underestimated. Current paper contributed towards the development of the numerical approaches in the marine sciences demonstrating the applied usage of Python. The use of Python proved to be an effective method for the aquatic or marine studies where thematic data has to be statistically processed and visualized. The code snippets or scripts in full are provided in the presented paper to enable repeatability of this research. Clear syntax and semantics of Python minimizes the chance of conflicts caused by hard-to-find bugs in a code.

The statistical data analysis has been implemented using Python machine learning libraries and algorithms. The seamless integration of the multi-dimensional data by the Matplotlib Python package enabled to perform marine geodata analysis and proper selection of the optimal algorithms to process the data features. The fully provided codes written on Python language make this research easily repeatable and accessible for other scientists. The adoption of the open source programming rapidly replacing the commercial technologies, Python will soon become popular among the marine geologists, geographers and Earth data analysts. This research contributes towards the development of the technical implementation of the machine learning algorithms for effective processing of the geodata.

Acknowledgments

This research was funded by the China Scholarship Council (CSC) State Oceanic Administration (SOA) Marine Scholarship of China, Grant Nr. 2016SOA002, Beijing, P.R.C.

References

Abdi, H. Williams, L. J. (2010). Principal component analysis. *Computational Statistics.* 2 (4): 433–459.

Bailey, RA. (2008). *Design of Comparative Experiments*. Cambridge University Press. ISBN 978-0-521-68357-9.

Bates DM., Watts, DG. (1988). *Nonlinear Regression Analysis and Its Applications.* Ed. by Kotlyakov, V.M. 2nd ed. New York: John Wiley & Sons, Inc. A Wiley-Interscience Publication, 102. ISBN: 0470139005.

Cartapanis, O., Jaccard, SL., Galbraith, E. (2018). Carbon burial in deep-sea sediment and implications for oceanic inventories of carbon and alkalinity over the last glacial cycle. *In: Climate of the Past Discussions.*

Contreras-Reyes, E., Jara, J., Maksymowicz, A., Weinrebe, W. (2013). Sediment loading at the southern Chilean Trench and its tectonic implications. *Journal of Geodynamics*, 66, 134–145.

Craglia, M., Jiri H., Stefano N., Mattia S.. (2017). Exploring the Depths of the Global Earth Observation System of Systems. *Big Earth Data 1* (1–2): 21–46.

Electronic copy available at: https://ssrn.com/abstract=3504388

Craglia, M., Kees de Bie, Davina J., Martino P., Gábor RF., Changlin W., Alessandro A., (2012). Digital Earth 2020: Towards the Vision for the Next Decade. *International Journal of Digital Earth* 5 (1): 4–21.

Dolotov, YS. (2010). Processes of Relief Formation and Sedimentation on the Tidal Coasts of the World Ocean. In Russian. Ed. by Safyanov, G.A. Moscow: *Scientific World*. ISBN: 978-5- 91522-235-8.

Dong, D., Zhang, Z., Bai, Y., Fan, J., Zhang, G. (2018). Topographic and sedimentary features in the Yap subduction zone and their implications for the Caroline Ridge subduction. *Tectonophysics*, 722, 410–421.

Dutkiewicz, A., Müller, R.D., O'Callaghan, S., Jónasson, H. (2015). Census of seafloor sediments in the worlds ocean. *Geology*, 43, 795–798.

Frigge, M., Hoaglin, D. C., Iglewicz, B. (1989). Some Implementations of the Boxplot. The American Statistician, 43(1): 50–54.

Gazioğlu C, Gökaşan, E, Algan O, Yücel ZY, Tok B, Doğan E (2002) Morphologic features of the Marmara Sea from multibeam data. *Mar Geol* 190:397–420

Gazioğlu C, Yucel ZY. Dogan E (2005) Morphological features of major submarine landslides of Sea of Marmara using multibeam data. *J Coastal Research*, 21:664–673.

Gelman, A. (2005). Analysis of variance? Why it is more important than ever. *The Annals of Statistics*, 33: 1–53.

Heuret, A., Conrad, C.P., Funiciello, F. (2012). Relation between subduction megathrust earthquakes, trench sediment thickness and upper plate strain. *Geophysical Research Letters*, 39(5), 131–138.

Huene, R. von, Scholl, D.W. (1991). Observations at Convergent Margins Concerning Sediment Subduction, Subduction Erosion, and the Growth of Continental Crust. *Review of Geophysics*, 29, 279–316.

Jiang, H.; Van Genderen, J.; Mazzetti, P.; Koo, H.; Chen, M. (2019). urrent status and future directions of geoportals.*Int. J. Digit. Earth*, 1-23.

Jimenez-Diaz, A., Ruiz, J., Pérez-Gussinyé, M., Kirby, J.F., Alvarez-Gomez, J.A., Tejero, R., Capote, R. (2014). Spatial variations of effective elastic thickness of the lithosphere in Central America and surrounding regions. *Earth Planetary Science Lettres*, 391, 55–66.

Jolliffe, IT. (2002) *Principal Component Analysis, Series: Springer Series in Statistics*, 2nd ed., Springer, NY, XXIX, 487, 28 illus. ISBN 978-0-387-95442-4

Judge, A.V., McNutt, M.K. (1991). The relationship between plate curvature and elastic plate thickness: a study of the Peru-Chile trench. Journal of Geophysical Research, 96 (B10), 16625–16639.

Kalnins, L.M., Watts, A.B. (2009). Spatial variations in effective elastic thickness in the Western Pacific Ocean and their implications for Mesozoic volcanism. *Earth and Planetary Science Letters*, 286 (1), 89–100.

Kaya, Ş., Müftüoğlu, O., Tüysüz, O. (2004). Tracing the geometry of an active fault using remote sensing and digital elevation model: Ganos Segment, North Anatolian Fault Zone, Turkey. *International Journal of Remote Sensing*, *25*, 3843–3855.

King, S.D., Masters, G. (1992). An inversion for radial viscosity structure using seismic tomography. *Geophysical Research Letters*, 19, 1551–1554.

Lacharité, M., Metaxas, A. (2017). Hard substrate in the deep ocean: How sediment features influence epibenthic megafauna on the eastern Canadian margin. *Deep-Sea Research Part* I, 126, 50–61.

Lehu, R., Lallemand, S., Hsu, S.-K., Babonneau, N., Ratzov, G., Lin, A.T., Dezileau, L. (2015). Deep-sea sedimentation offshore eastern Taiwan: Facies and processes characterization. *Marine Geology*, 369, 1–18.

Lisicyn, A.P. (1974). *Osadkoobrasovanie v okeanah (Sedimentation in the oceans).* In Russian. Ed. by Bezrukov, P.L. Moscow: Nauka.

McGill, R., Tukey, JW., Larsen, WA. (1978). Variations of Box Plots. *The American Statistician*, 32 (1): 12–16. doi:10.2307/2683468.

Mckenzie, D., Bowin, C. (1976). Relationship between bathymetry and gravity in Atlantic Ocean. *Journal of Geophysical Research*, 81 (11), 1903–1915.

Montgomery, DC. (2001). *Design and Analysis of Experiments.* 5th ed. New York: Wiley. ISBN 978-0-471-31649-7.

Ricard, Y., Vigny, C., Froidevaux, J. (1989). Mantle heterogeneities, geoid and plate motion: a MonteCarlo inversion. *Journal of Geophysical Research*, 94, 13739–13754.

Sá, JPM. de. (2007). *Applied Statistics Using SPSS, Statistica, Matlab and R.* 2nd ed. Library of Congress Control Number: 2007926024. Porto, Portugal: Springer, 520 ISBN: 978- 3-540-71971-7

Sandwell, DT. Smith, WHF (2009). Global marine gravity from retracked Geosat and ERS -1 altimetry: ridge segmentation versus spreading rate. *Journal of Geophysical Research*, 114, 1-18.

Sandwell, DT., Smith, WHF. (1997). Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *Journal of Geophysical Research*, 102 (B5), 10,039-10,054.

Sattarova, VV., Aksentov, KI. (2018). Geochemistry of mercury in surface sediments of the Kuril Basin of the Sea of Okhotsk, Kuril-Kamchatka Trench and adjacent abyssal plain and northwest part of the Bering Sea. *Deep-Sea Research Part* II, 154, 24–31.

Smith WHF, Sandwell DT. (1997) Global sea floor topography from satellite altimetry and ship depth soundings, *Science*, 277, 1956–62.

Turnewitsch, R., Falahat, S., Stehlikova, J., Oguri, K., Glud, R.N., Middelboe, M., Kitazato, H., Wenzhöfer, F., Ando, K., Fujio, S., Yanagimoto, D. (2014). Recent sediment dynamics in hadal trenches: Evidence for the influence of higher-frequency (tidal, near-inertial) fluid dynamics. *Deep- Sea Research I*, 90, 125–138.

Udintsev, G.B. (1987). *Relief i stroenie dna okeanov (Relief and the structure of the ocean floor).* In Russian. Ed. by Il'yin, A.V. Moscow: Nedra.

Völker, D., Reichel, T., Wiedicke, M., Heubeck, C. (2008). Turbidites deposited on Southern Central

Electronic copy available at: https://ssrn.com/abstract=3504388

Chilean seamounts: Evidence for energetic turbidity currents. *Marine Geology*, 251, 15–31.

Wessel, P., Smith, W.H.F. (1995) *New version of the Generic Mapping Tools released*, EOS Trans. AGU, 76, 329.

Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J. Wobbe, F. (2017) GMT Documentation Release 5.4.2. 314.

Yang, A., Fu, Y. (2018). Estimates of effective elastic thickness at subduction zones. *Journal of Geodynamics*, 117, 75–87.

Yang, J., Cui, Z., Dada, O.A., Yang, Y., Yu, H., Xu, Y., Lin, Z., Chen, Y. Tang, X.(2018). Distribution and enrichment of trace metals in surface marine sediments collected by the manned submersible Jiaolong in the Yap Trench, northwest Pacific Ocean. *Marine Pollution Bulletin*, 135, 1035–1041.

Zheng, H.-W., Gao, R., Li, T.-D., Li, Q.-S., He, R.-W. (2013). Collisional tectonics between the Eurasian and Philippine Sea plates from tomography evidences in Southeast China. *Tectonophysics*, 606, 14–23.

Zhou, Z., & Lin, J. (2018). Elastoplastic deformation and plate weakening due to normal faulting in the subducting plate along the Mariana Trench. Tectonophysics, 734–735, 59–68

Zhou, Z., Lin, J., & Behn, M.D. (2015). Mechanism for normal faulting in the subducting plate at the Mariana Trench. *Geophysical Research Letters*, 42, 4309–4317.

Zhu, G., Shi, Y., Tackley, P. (2010). Subduction of the Western Pacific Plate underneath Northeast China: Implications of numerical studies. *Physics of the Earth and Planetary Interiors*, 178, 92–99.

Electronic copy available at: https://ssrn.com/abstract=3504388